Exploratory Data Analysis on Zomato Restaurants

By Ahmed Saghir

Introduction

Zomato is an online platform allowing customers to order food from various restaurants. The dataset we have show's information related to restaurants, including their names, locations, cuisines, price ranges, ratings, and much more.

Objective

To utilize exploratory data analysis (EDA) skills to understand customer preferences, dining trends, and competitive landscape in various regions of India, and to design an effective marketing campaign for a restaurant chain.

Dataset: "zomato_restaurants_in_India.csv"

Data Overview

The data set contains total 211944 rows and 26 columns.

res_id (int64): Unique ID for each restaurant.
name (object): Name of the restaurant.
establishment (object): Type of establishment (e.g., Quick Bites, Cafe).
url (object): URL link to the Zomato restaurant page.
address (object): Complete address of the restaurant.
city (object): City where the restaurant is located.
city_id (int64): Unique ID for each city
locality (object): Locality within the city.
latitude (float64): Latitude coordinate of the restaurant.
longitude float64): Longitude coordinate of the restaurant.
zipcode (object): Zipcode of the restaurant's location
country_id (int64): Unique ID for country (India)
locality_verbose (object): Description of the locality.
cuisines (object): Types of cuisines offered by restaurants.
timings (object): Operating hours of the restaurant
average_cost_for_two (int64): Average cost for two people.
price_range (int64): price range category (likely from 1 to 4).
currency (object): Currency used for pricing (Indian Rupee).
highlights (object): Special features or popular services of the restaurant
aggregate_rating (float64): Aggregated rating of the restaurant.
rating_text (object): Descriptive rating of the restaurant. (e.g., "Very Good", "Excellent").
votes (int64): Number of votes for the restaurant.
photo_count (int64): Number of photos posted for the restaurant.
opentable_support (float64): Indicates for OpenTable support.
delivery (int64): Indicates if delivery service is available.
takeaway (int64): Indicates if takeaway is available.

Data Cleaning and Preparation

Check Missing Values

address: 134 missing values.

zipcode: 163,187 missing values, a significant number.

cuisines: 1,391 missing values.

timings: 3,874 missing values.

Opentable_support: 48 missing values.

Handling Duplicated Values

The dataset contains 151,527 duplicated rows. After removing the duplicated rows, we have 60,417 rows and 26 columns.

Handling Null Values

Removed 'zipcode': This column was removed because **76.9**% of the values are missing.

Removed 'url': This column contained url link which is useless for our analysis.

Removed 'opentable_support': This column was removed because it contained either 0 and nan values.

Removed 'country_id': This column was removed because it only contained 1 unique value ('1').

Removed 'currency': This column was removed because it only contained 1 unique value ('Rs.').

Handling missing values in 'address', 'cuisines' and 'timings' columns:

After removing duplicated rows, **address has 18 missing** values, **cuisines have 470 missing** values and **timings has 1070 missing** values.

For timings column: took mode value and replaced the missing values with this value

(11 AM to 11 PM).

For cuisines column: missing values were replaced with **"Unknown".**

For 'address' column: By using 'geopy' library, we imputed the missing values with the help of 'latitude' and 'longitude' column.

Feature Engineering

After handling missing values, the dataset contains **60417 observations** and **21 features**.

For better readability, renamed **"average_cost_for_two**" to **"cost2plates"**.

Found empty brackets in **'establishment'** column (['Quick Bites'], ['Bakery'], ['']). replacing it with Not Specified and removed square brackets and quotes from the column (Quick Bites, Bakery, Not Specified).

Found **text_rating** in other languages like (Dobre, Baik). As the occurrence of this values were less, hence created a group name **'others'** and added all these entries into it (Excellent, Poor, others).

Separated '**highlights**' by comma (,) for better analysis.

Exploratory Data Analysis

Descriptive Statistics

The descriptive statistics for Numerical Columns are as follows:

	count	mean	std	min	25%	50%	75%	max
res_id	60417.0	1.309335e+07	8.132809e+06	50.0	3.000488e+06	1.869150e+07	1.886666e+07	1.915979e+07
city_id	60417.0	3.418302e+03	5.179352e+03	1.0	7.000000e+00	2.600000e+01	1.129500e+04	1.135400e+04
latitude	60417.0	2.134943e+01	4.118800e+01	0.0	1.632476e+01	2.232088e+01	2.674439e+01	1.000000e+04
longitude	60417.0	7.658804e+01	1.060051e+01	0.0	7.465403e+01	7.713531e+01	7.992819e+01	9.183277e+01
cost2plates	60417.0	5.383045e+02	5.938522e+02	0.0	2.000000e+02	4.000000e+02	6.000000e+02	3.000000e+04
price_range	60417.0	1.730821e+00	8.804625e-01	1.0	1.000000e+00	1.000000e+00	2.000000e+00	4.000000e+00
aggregate_rating	60417.0	3.032868e+00	1.440751e+00	0.0	2.900000e+00	3.500000e+00	4.000000e+00	4.900000e+00
votes	60417.0	2.615749e+02	7.282842e+02	-18.0	7.000000e+00	4.200000e+01	2.070000e+02	4.253900e+04
photo_count	60417.0	1.942474e+02	7.056825e+02	0.0	1.000000e+00	1.100000e+01	8.200000e+01	1.770200e+04
delivery	60417.0	-3.717993e-01	9.252490e-01	-1.0	-1.000000e+00	-1.000000e+00	1.000000e+00	1.000000e+00
takeaway	60417.0	-1.000000e+00	0.000000e+00	-1.0	-1.000000e+00	-1.000000e+00	-1.000000e+00	-1.000000e+00

The descriptive statistics for Categorical Columns are as follows:

	count	unique	top	freq
name	60417	41100	Domino's Pizza	406
establishment	60417	27	Quick Bites	15477
address	60417	50658	Laxman Jhula, Tapovan, Rishikesh	37
city	60417	99	Chennai	2612
locality	60417	3731	Civil Lines	804
locality_verbose	60417	3910	Gomti Nagar, Lucknow	315
cuisines	60417	9383	North Indian	4587
timings	60417	7740	11 AM to 11 PM	8748
highlights	60417	31455	['Dinner', 'Takeaway Available', 'Lunch', 'Cas...	925
rating_text	60417	7	Good	17569

Skewness:

Aggregate Rating Skewness: -1.3807068750066696
Votes Skewness: 12.764799913007979
Photo Count Skewness: 9.094059660523177
Average Cost of 2 Skewness: 6.055282592508311
Price Range Skewness: 0.9963448926604421

Central Tendency and Dispersion:

aggregate_rating: The average rating is around 3.03 with a standard deviation of 1.44, indicating variability in how restaurants are rated. Most ratings cluster around the median of 3.5.
votes: Restaurants have an average of 261 votes, but with a high standard deviation of 728, showing that some restaurants have many more votes than others.

Distribution Shape:

The skewness of key numerical fields:
aggregate_rating has a negative skew (-1.38), suggesting a concentration of ratings at the higher end.
votes and photo_count are highly positively skewed (12.76 and 9.09 respectively), indicating a long tail on the right side where few restaurants have very high votes and photo counts.
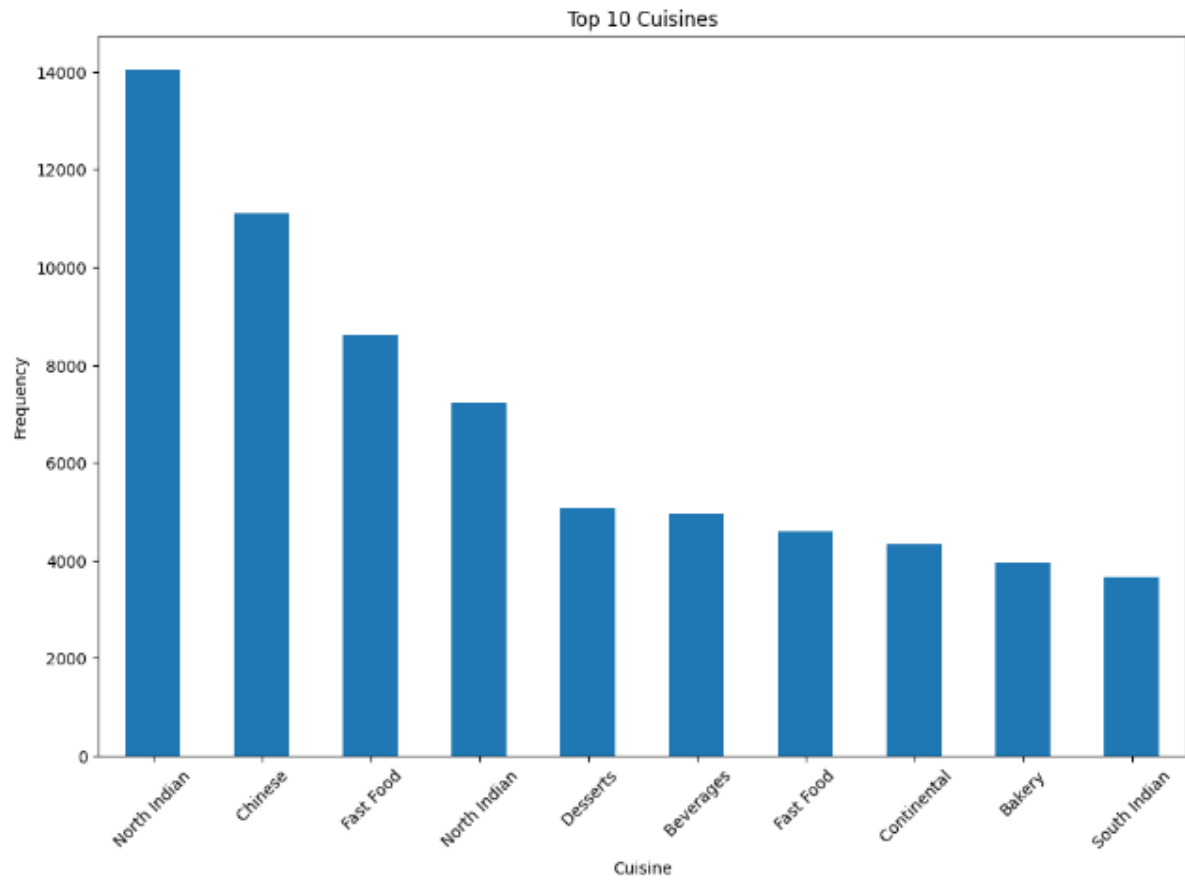
Categorical Data:

establishment has 27 unique types with 'Quick Bites' being the most common.
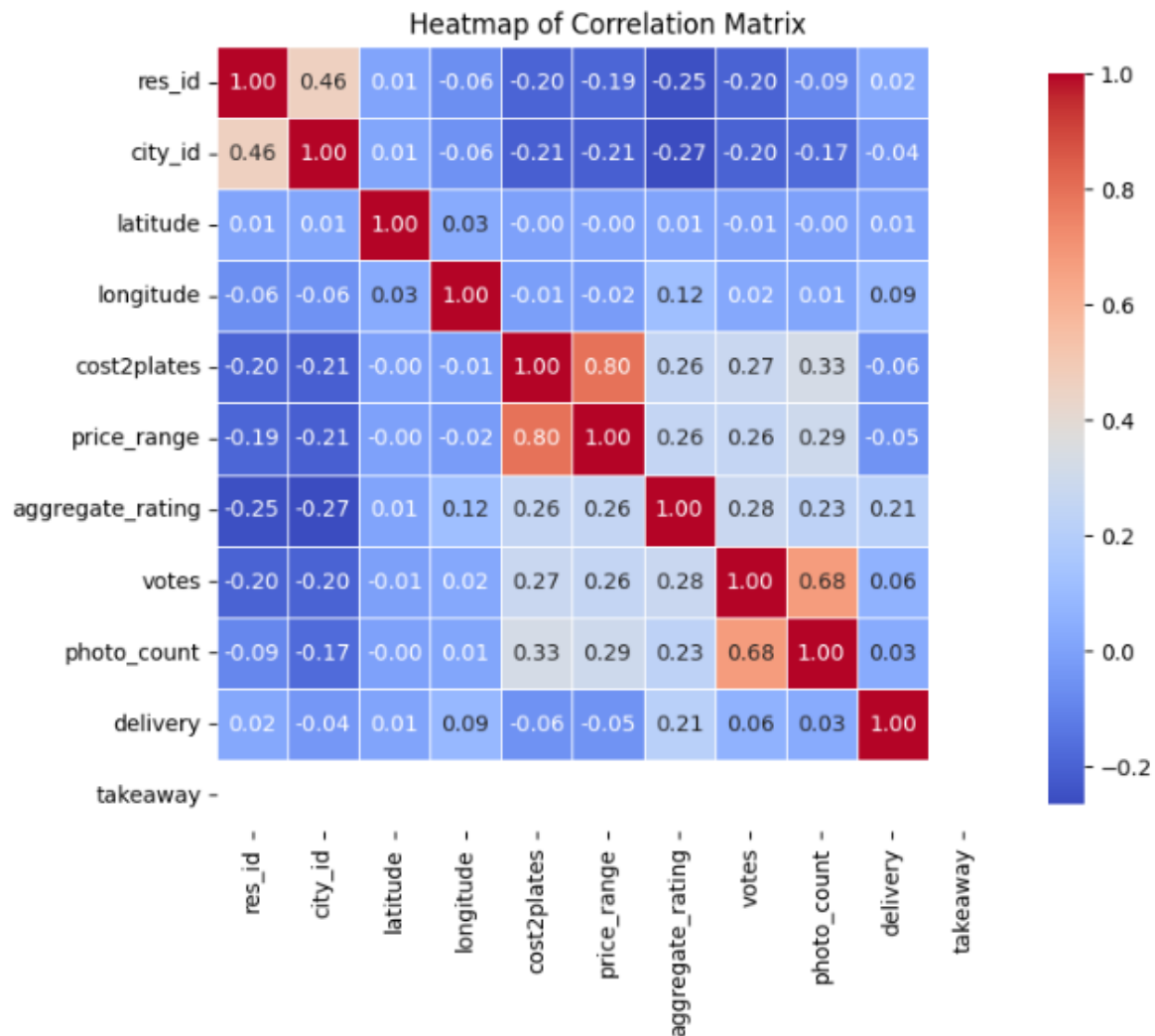**city:** The dataset includes 99 unique cities, with Chennai appearing most frequently.
rating_text: 'Good' is the most frequent rating.

Data Distribution Analysis

Distribution of Aggregate Ratings

Distribution of Price Ranges

Top 10 Cuisines

Distribution of Establishment
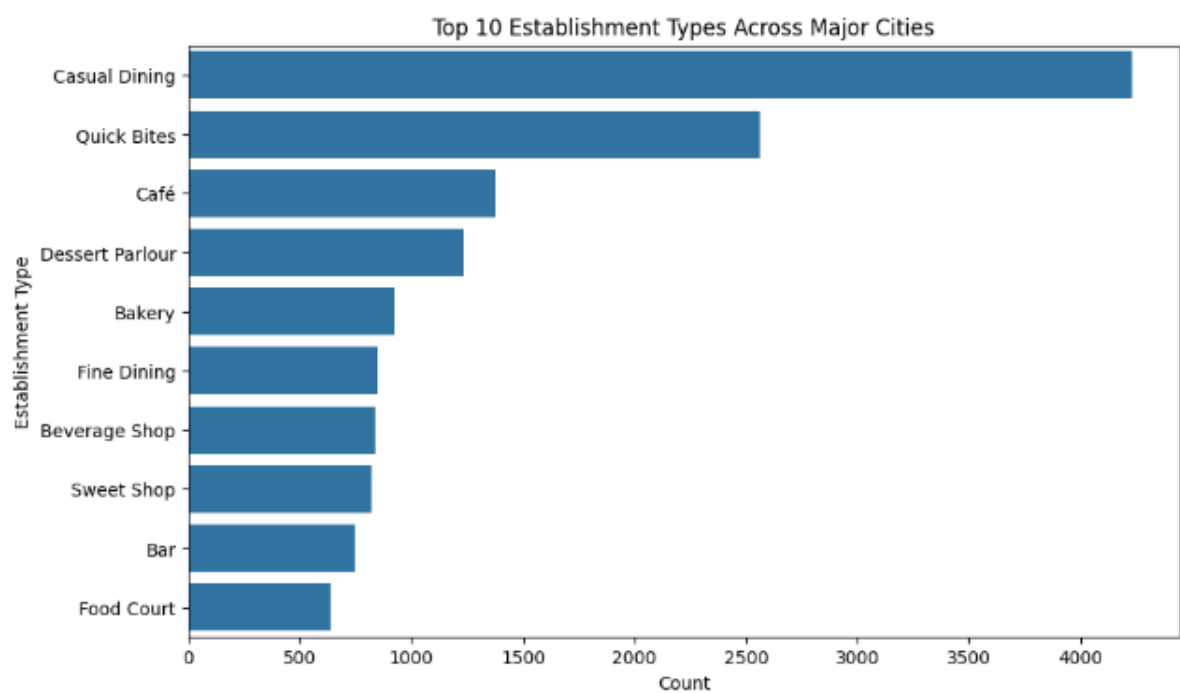
Correlation Analysis

Heatmap of Correlation Matrix
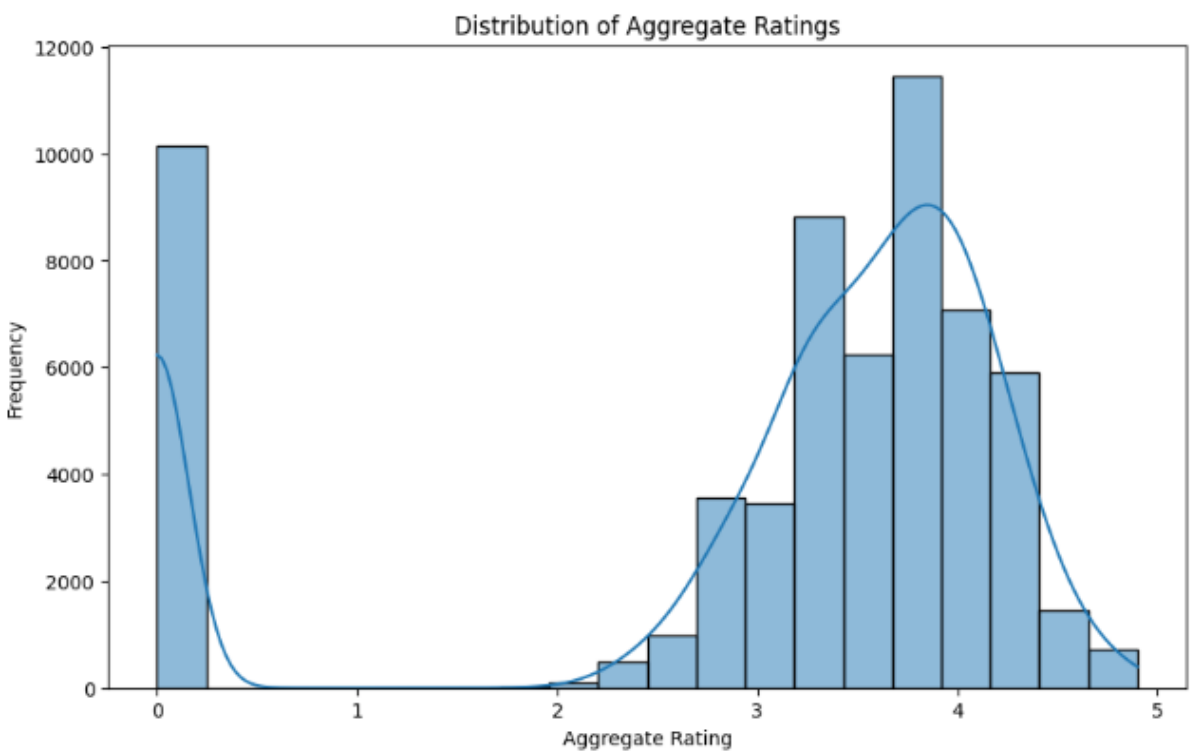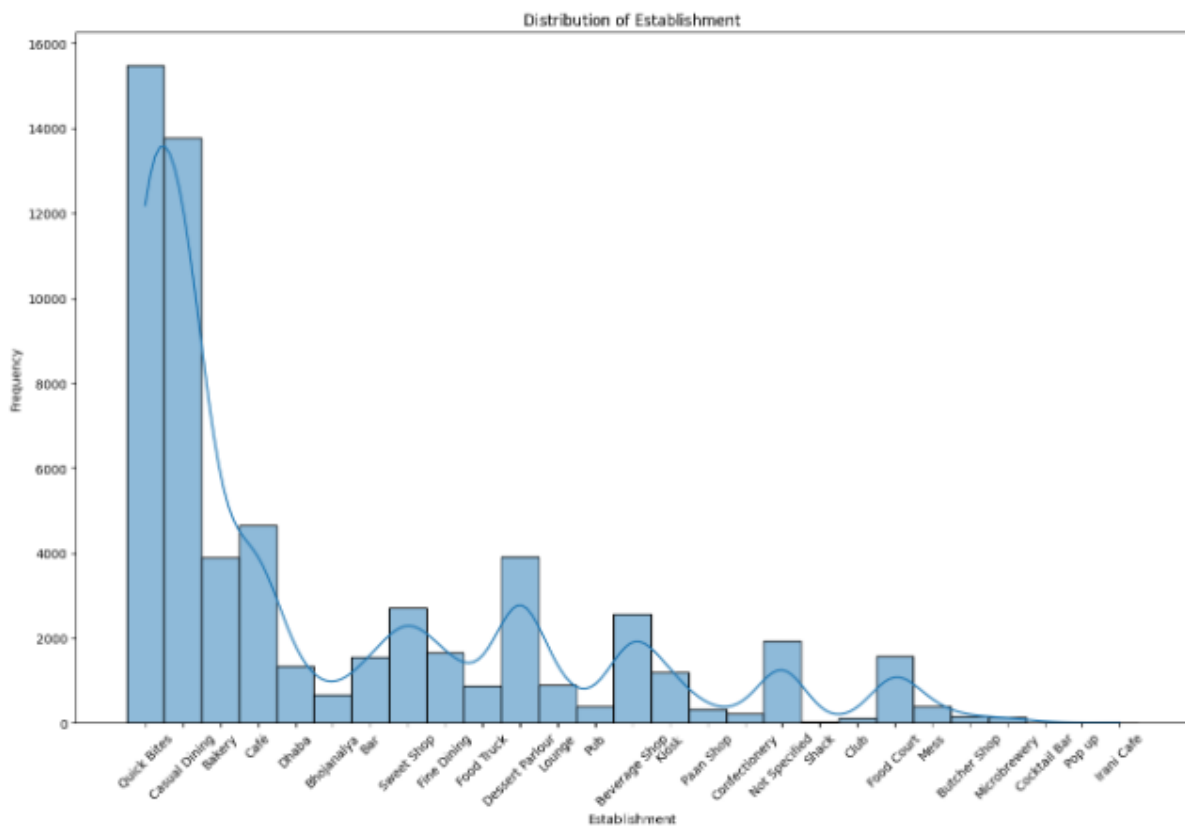
- **Average cost for two and price range:** There is a strong positive correlation (approximately 0.80) between these two variables, suggesting that as the average cost for two increases, the price range also tends to increase. This is expected as both are measures of cost.

- **Votes and photo _count:** There is also a strong positive correlation (approximately 0.68) between the number of votes a restaurant has and the number of photos posted. This indicates that more popular restaurants (with more votes) tend to have more photos posted by users.
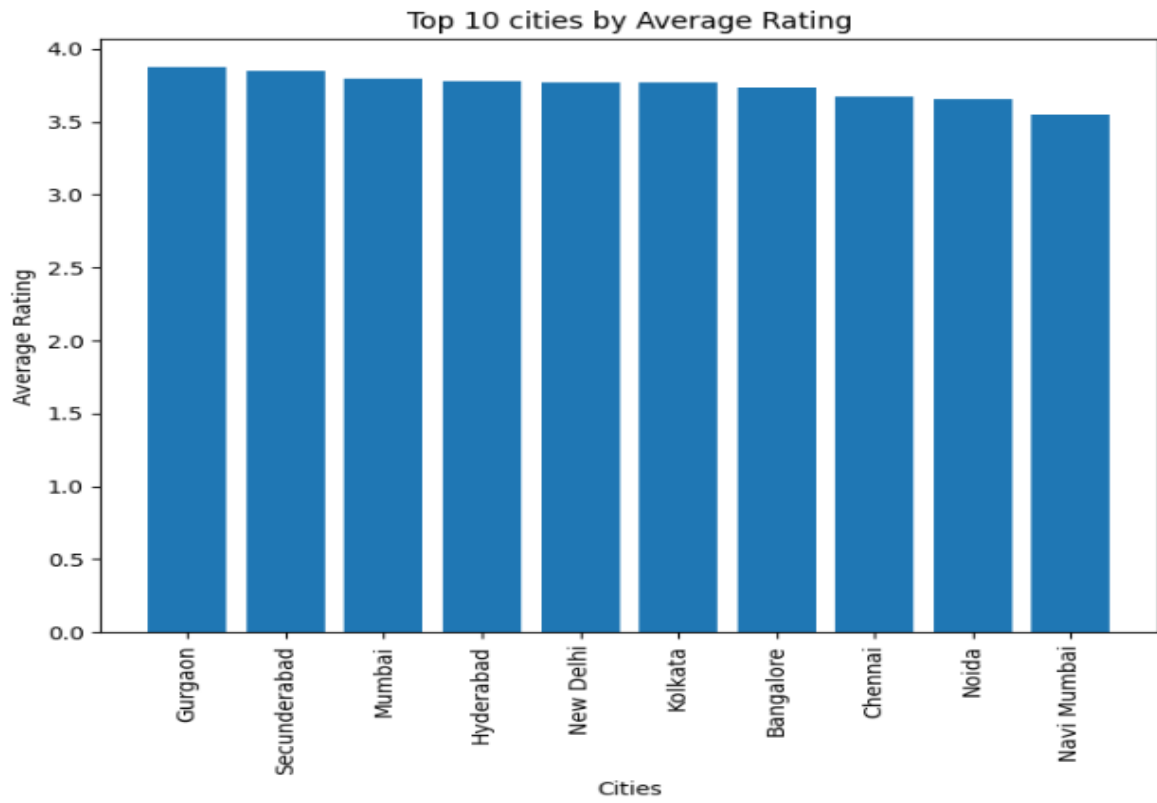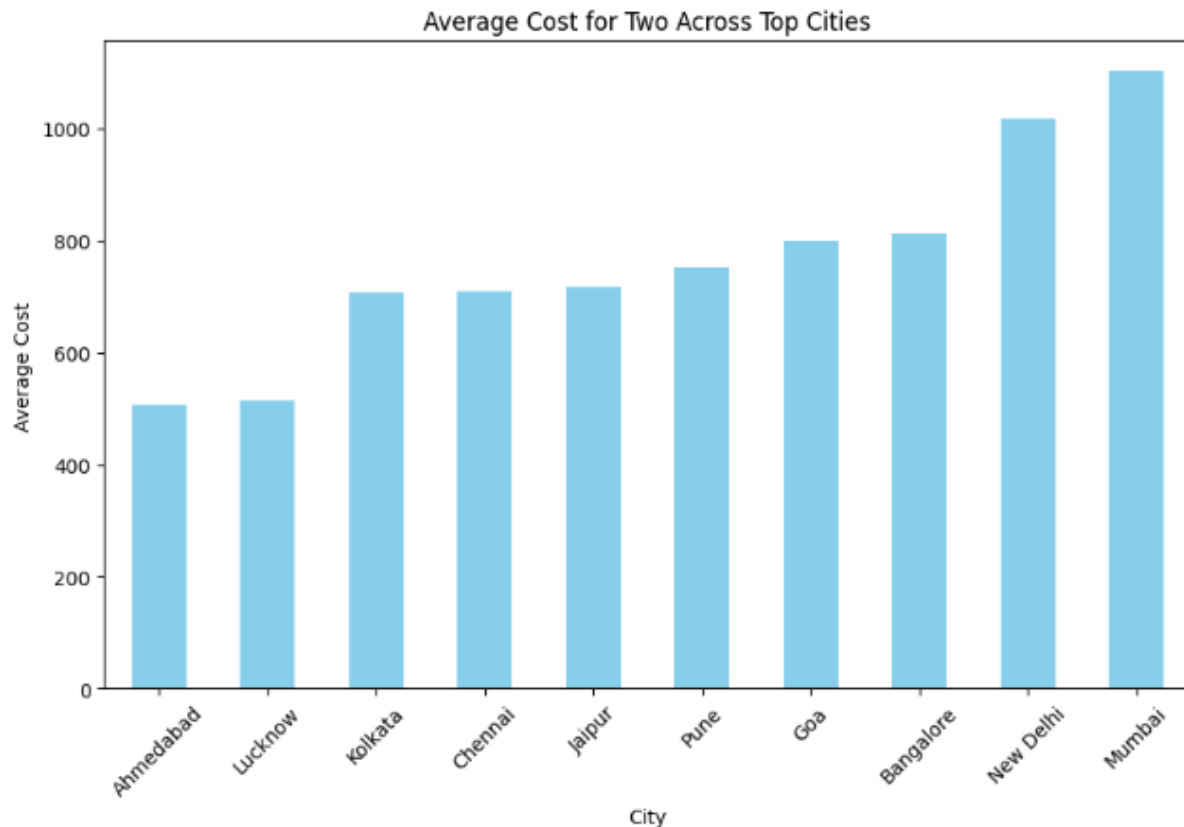
Regional Analysis

Restaurants Trends and Customer Preferences

Top 10 cities with the most listings are as following:

1. Chennai: 2612 Restaurants
2. Mumbai: 2538 Restaurants
3. Bangalore: 2365 Restaurants
4. Pune: 1911 Restaurants
5. New Delhi: 1847 Restaurants
6. Jaipur: 1456 Restaurants
7. Kolkata: 1413 Restaurants
8. Ahmedabad: 1329 Restaurants
9. Lucknow: 1290 Restaurants
10. Goa: 1169 Restaurants

Top 10 Establishment Types Across Major Cities

Top 10 cities by Average Rating

Average Price Range Across Top Cities

Average Cost for Two Across Top Cities
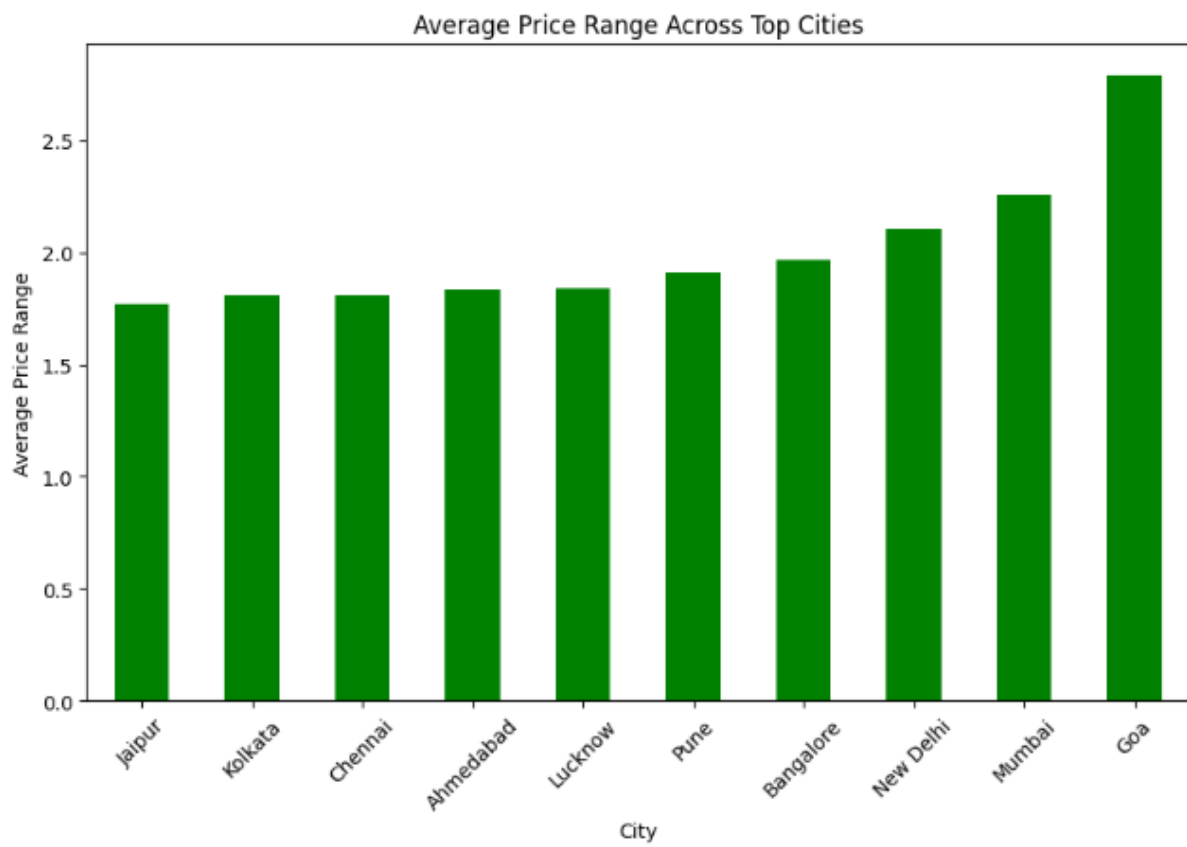
- Casual Dining and Quick Bites are the most common types of establishments across these major cities. Cafes, Dessert Parlours, and Bakeries also appear frequently.
- Mumbai has the highest average cost for two people dining, followed by New Delhi and Bangalore, indicating these cities might cater to a more upscale market.
  On the more affordable side, we have Ahmedabad and Lucknow, which have lower average dining costs.
- Goa stands out with a higher average price range, possibly due to its status as a tourist destination.
  Mumbai and New Delhi also show higher price ranges, reflecting their metro status and potentially higher living costs.
  Cities like Chennai, Kolkata, and Jaipur maintain a more moderate pricing profile.

Dining Preferences

Top 5 Popular Cuisines in Each City
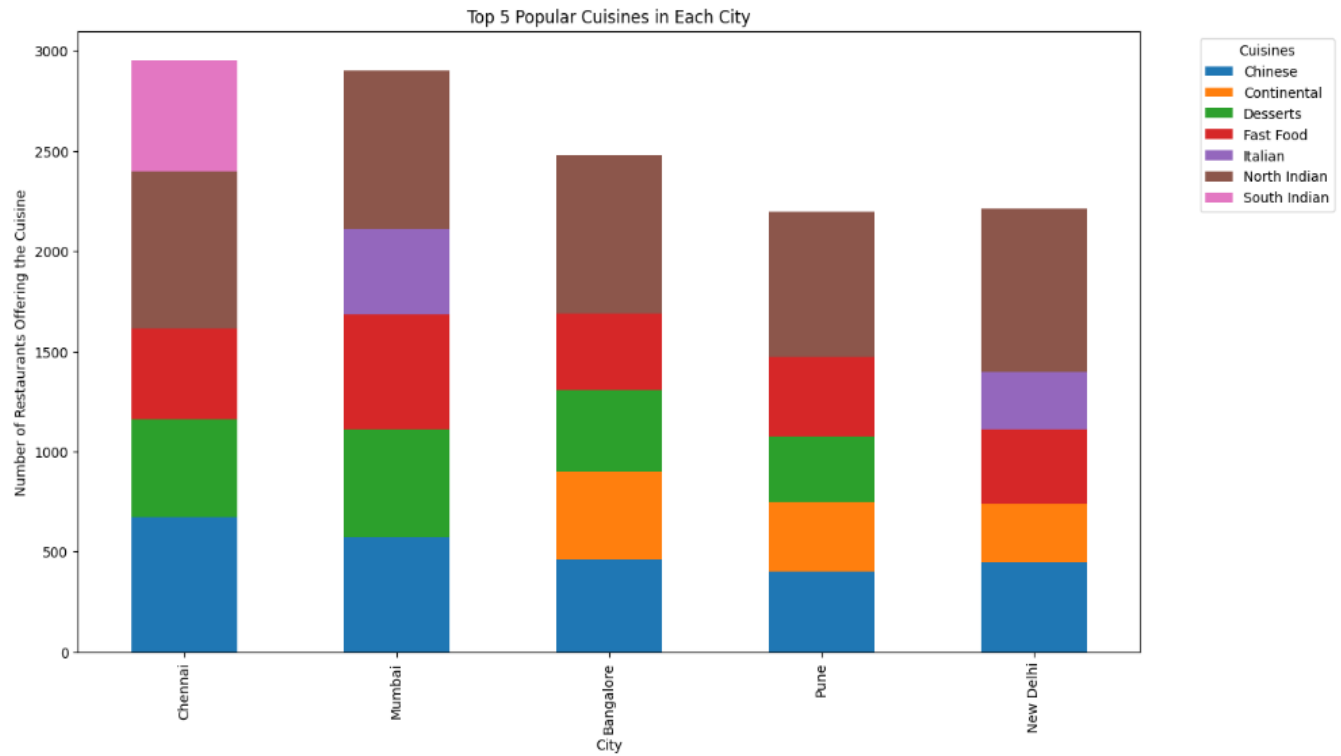
The top three cuisines for each of the major cities

- **Chennai:** North Indian, Chinese, South Indian

- **Mumbai:** North Indian, Chinese, Fast Food

- **Bangalore:** North Indian, Chinese, Continental

- **Pune:** North Indian, Chinese, Fast Food

- **New Delhi:** North Indian, Chinese, Fast Food

Rating Across Top Cities
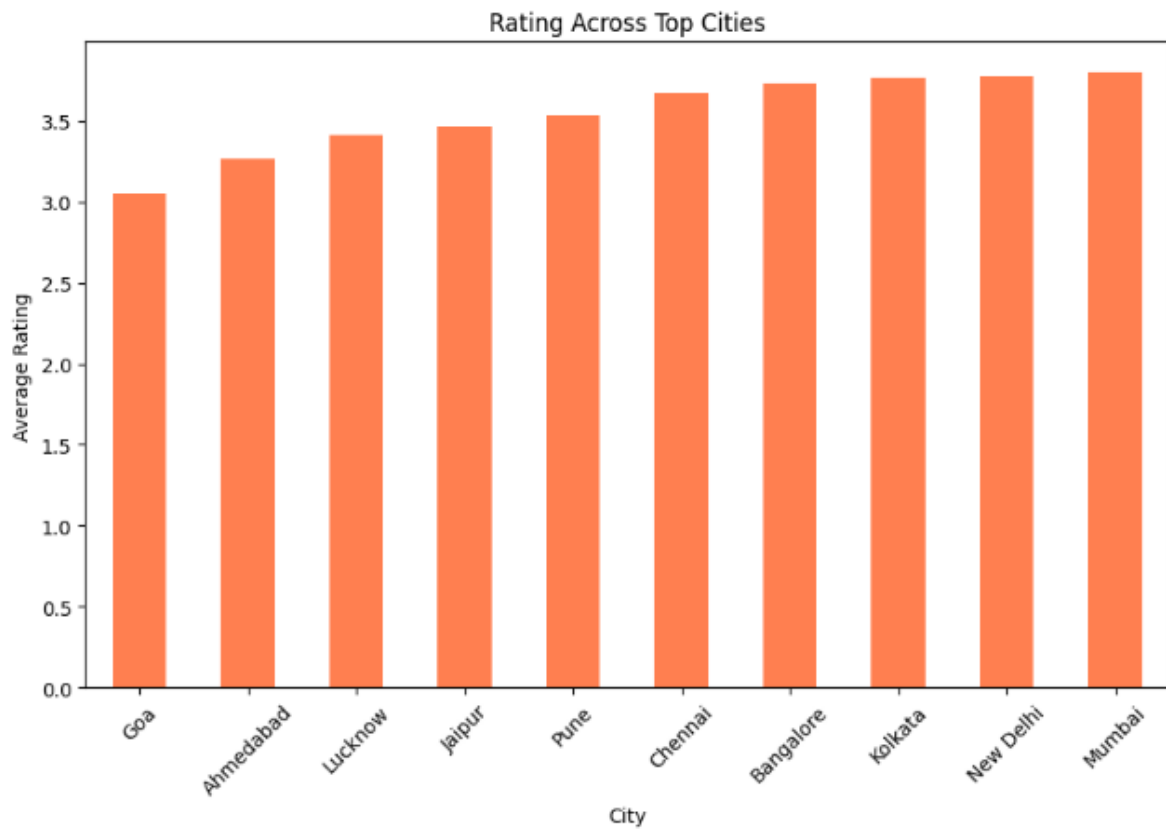
- Kolkata, New Delhi and Mumbai exhibit the highest average customer ratings, suggesting a high level of customer satisfaction or possibly better-quality restaurants

- Goa has a notably lower average rating, which could reflect a variety of factors, including tourist- services variations.

Customer Preference Analysis

Top Choice in Cuisines


Top 5 Cuisines in Selected Cities
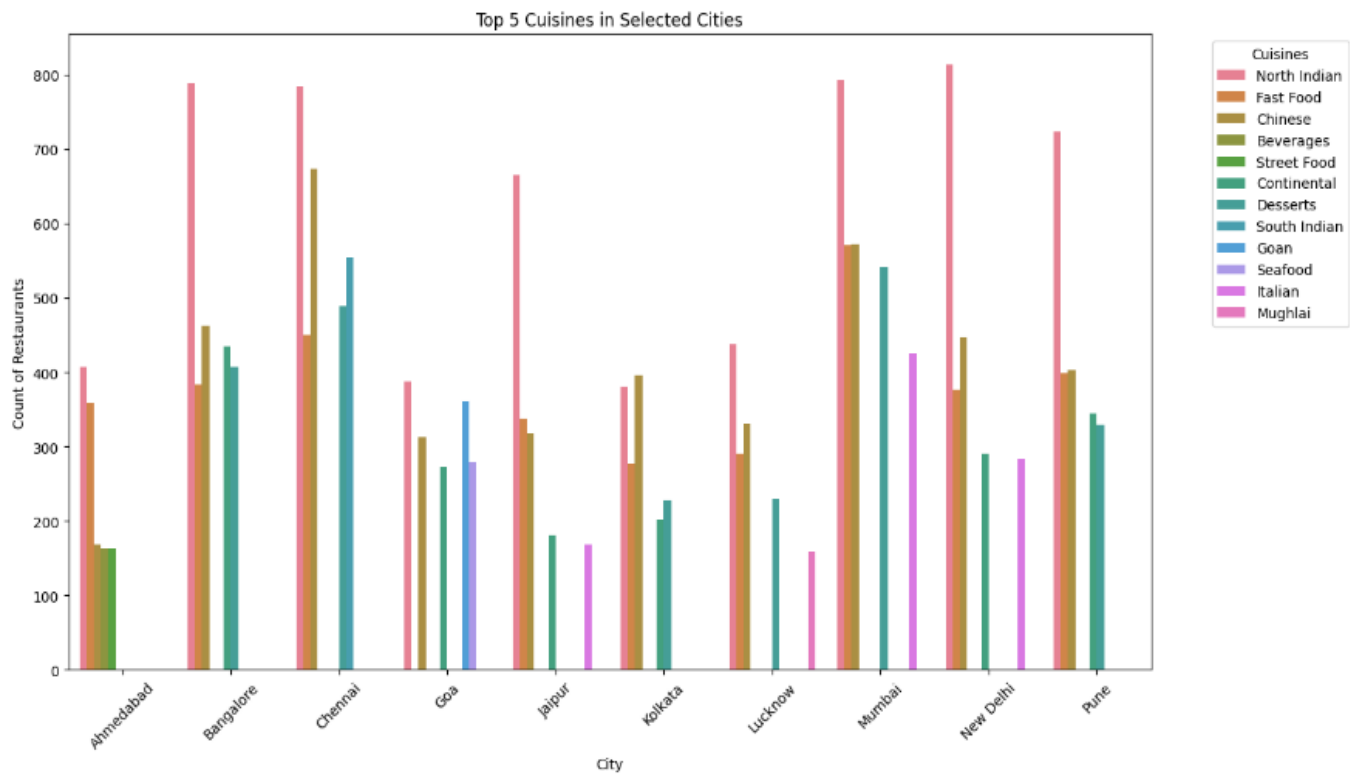
- **Varied Preferences:** While North Indian cuisine remains popular across these cities, we can see a diverse range of other cuisines featuring prominently. For instance, Fast Food and Chinese are common across all cities.
- **Regional Influences:** Cities like Chennai and Kolkata show a strong preference for local cuisines such as South Indian in Chennai and Bengali in Kolkata, reflecting regional culinary traditions.
- **International Flavours:** Cities like Mumbai and Pune, known for their cosmopolitan populace, display a wider variety of cuisines including Italian and Continental, indicating a palate for international flavours.
- **Unique Trends:** Each city has its own unique set of popular cuisines which could be influenced by local tastes, demographics, and cultural factors. For example, Street Food is notably popular in Ahmedabad.

Popularity by Customers

Scatter Plot of Aggregate Rating vs Votes
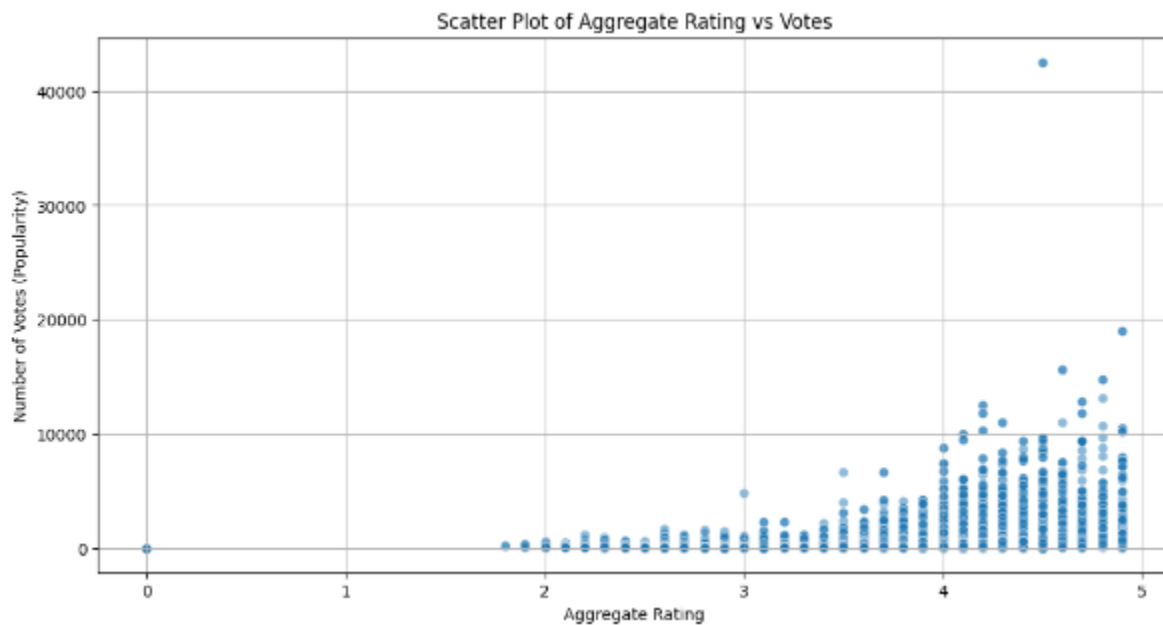
The scatter plot of aggregate ratings versus the number of votes shows a trend where higher-rated restaurants tend to receive more votes.

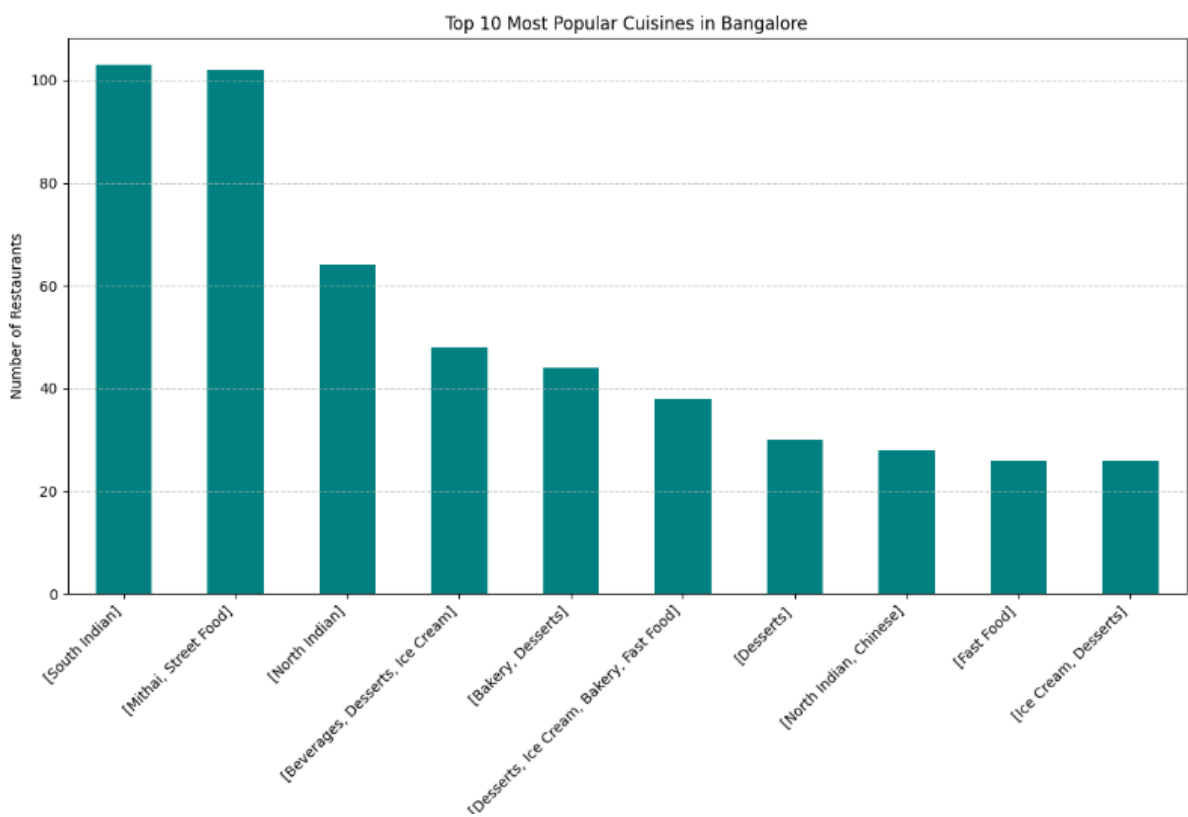Box Plot of Aggregate Rating across Price Ranges
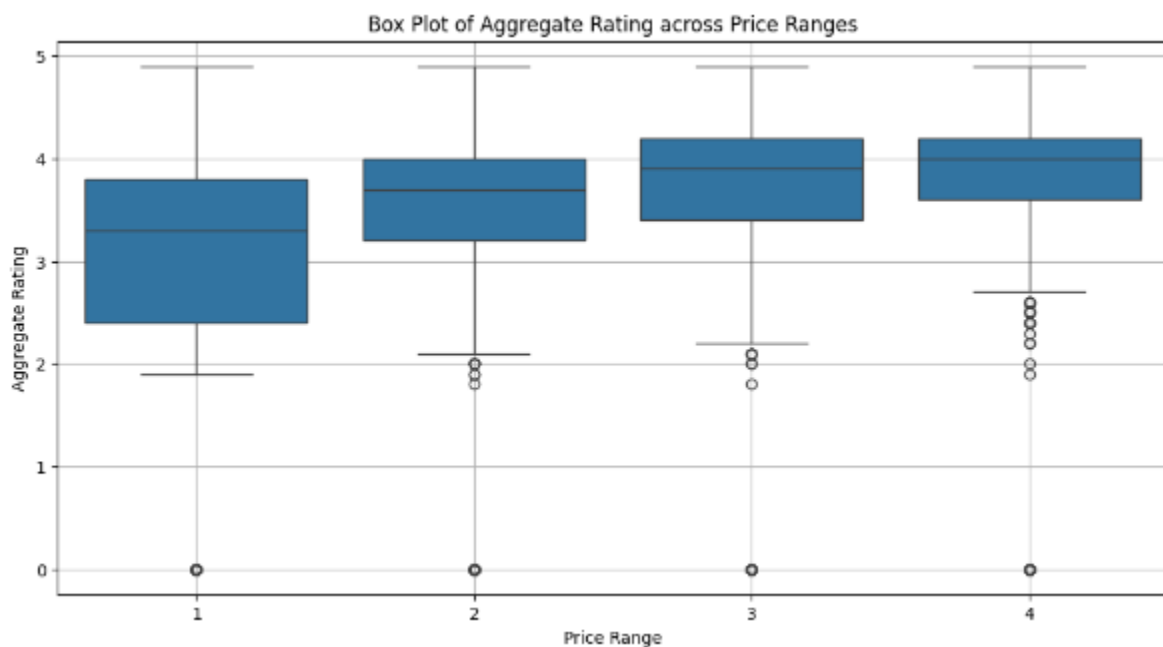
Higher price ranges tend to have a slightly higher median rating. This indicates that more expensive restaurants generally receive better ratings.
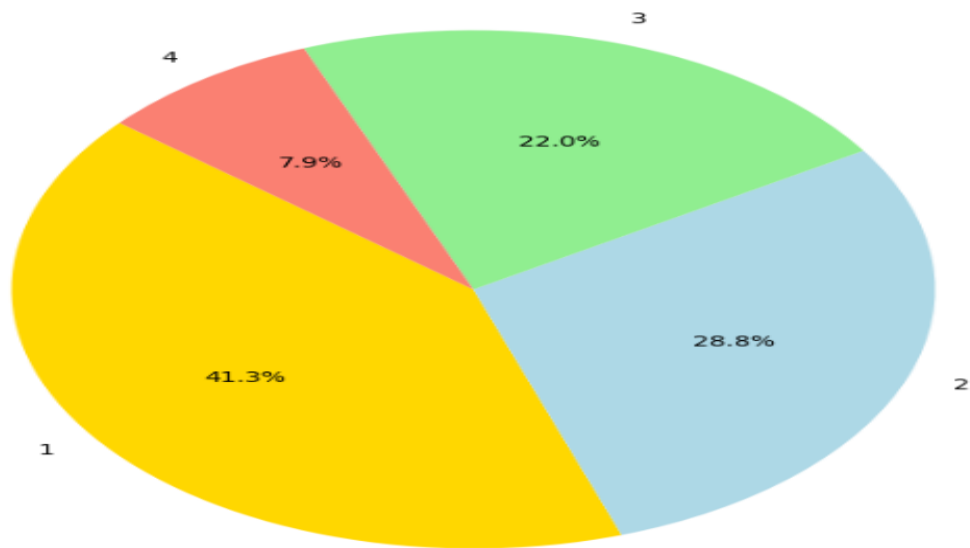
Competitive Analysis

Major Competitors

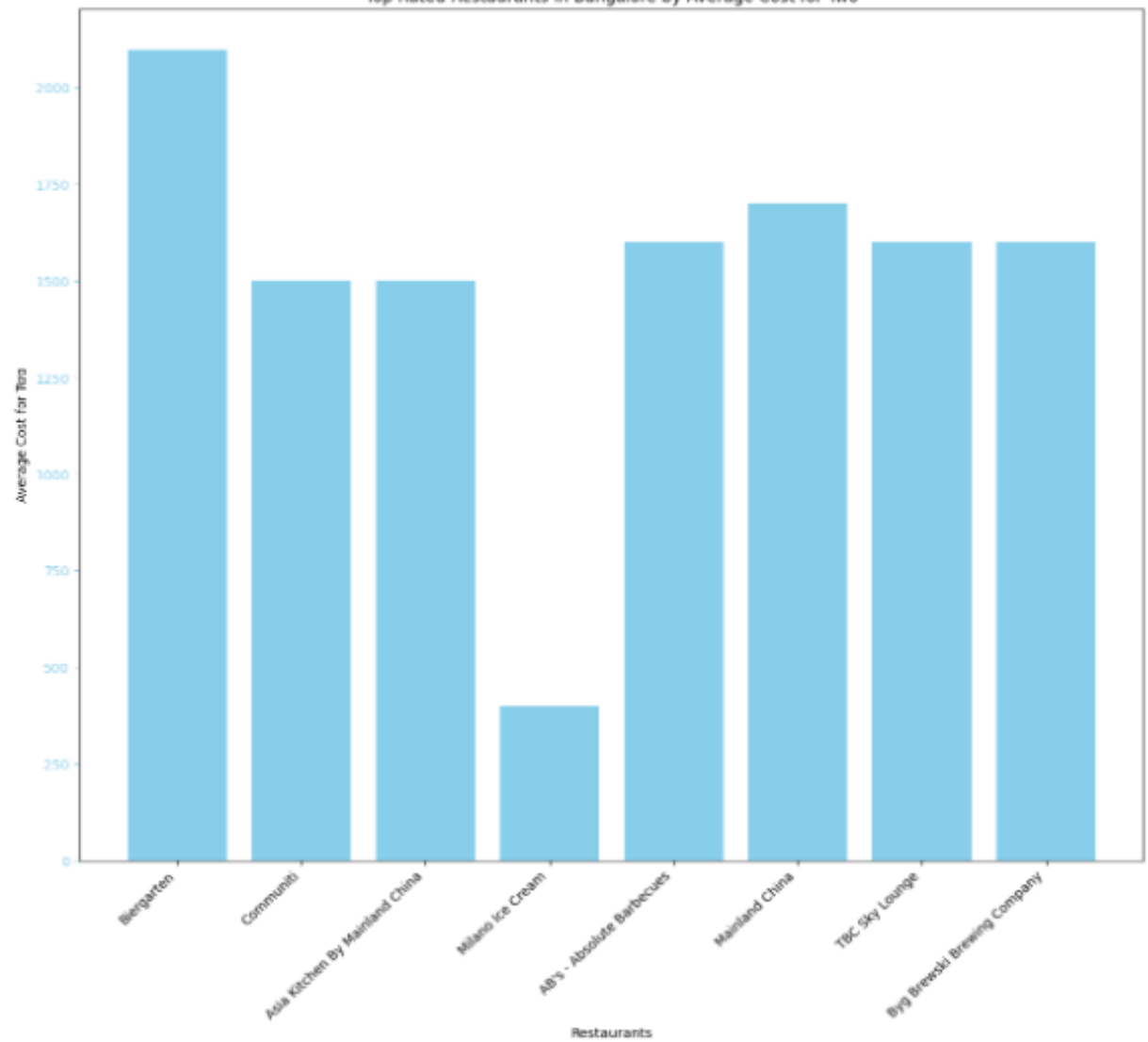Selecting Top 3 Metropolitan Cities: Bangalore, Mumbai, and Chennai.

Bangalore Market Analysis

Price Range Distribution in Bangalore Restaurants

Top Rated Restaurants in Bangalore by Average Cost for Two

Cuisine Popularity:

- Most Popular Cuisines: The top cuisines in Bangalore include South Indian, North Indian, and a variety of dessert-related options such as Mithai, Street Food, and Ice Cream.
- Diverse Offerings: There's a strong representation of both local (South Indian) and versatile options like North Indian, Chinese, and Continental.

Pricing Analysis:

- Average Cost for Two: The average cost for dining out in Bangalore is approximately ₹812.
- Price Range Distribution: Most restaurants fall into the lower to mid pricing brackets, with:
- Price Range 1 (most affordable): 977 restaurants
- Price Range 2: 681 restaurants
- Price Range 3 (more premium): 520 restaurants
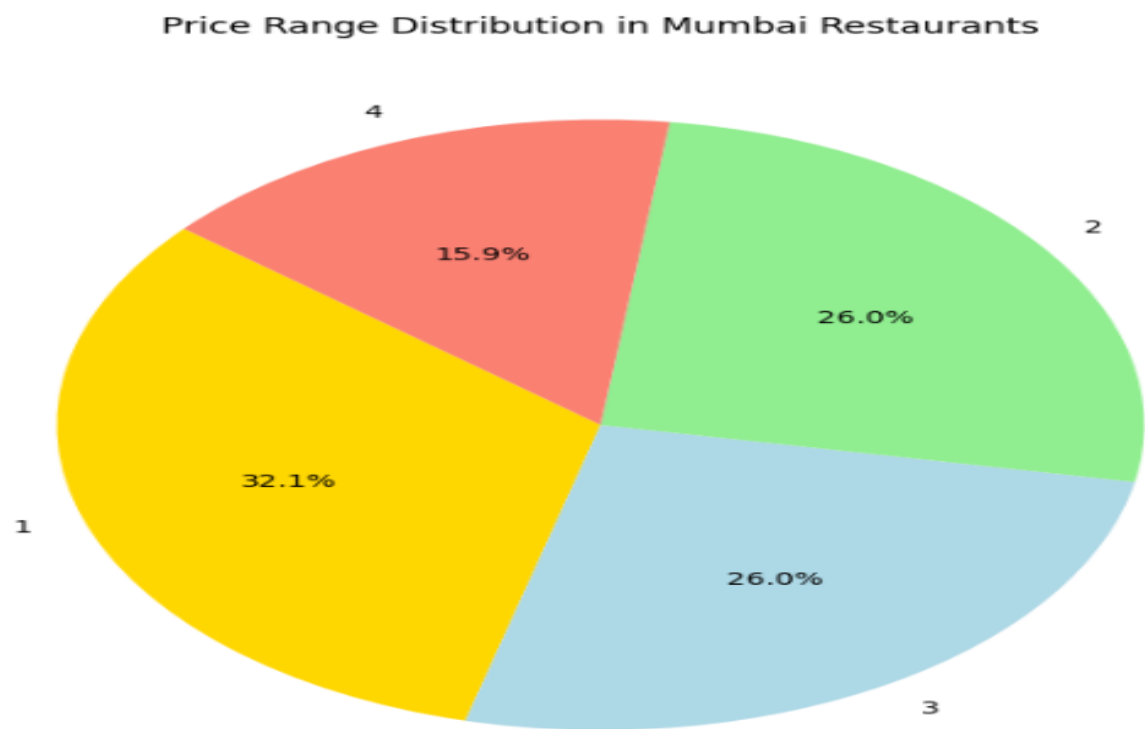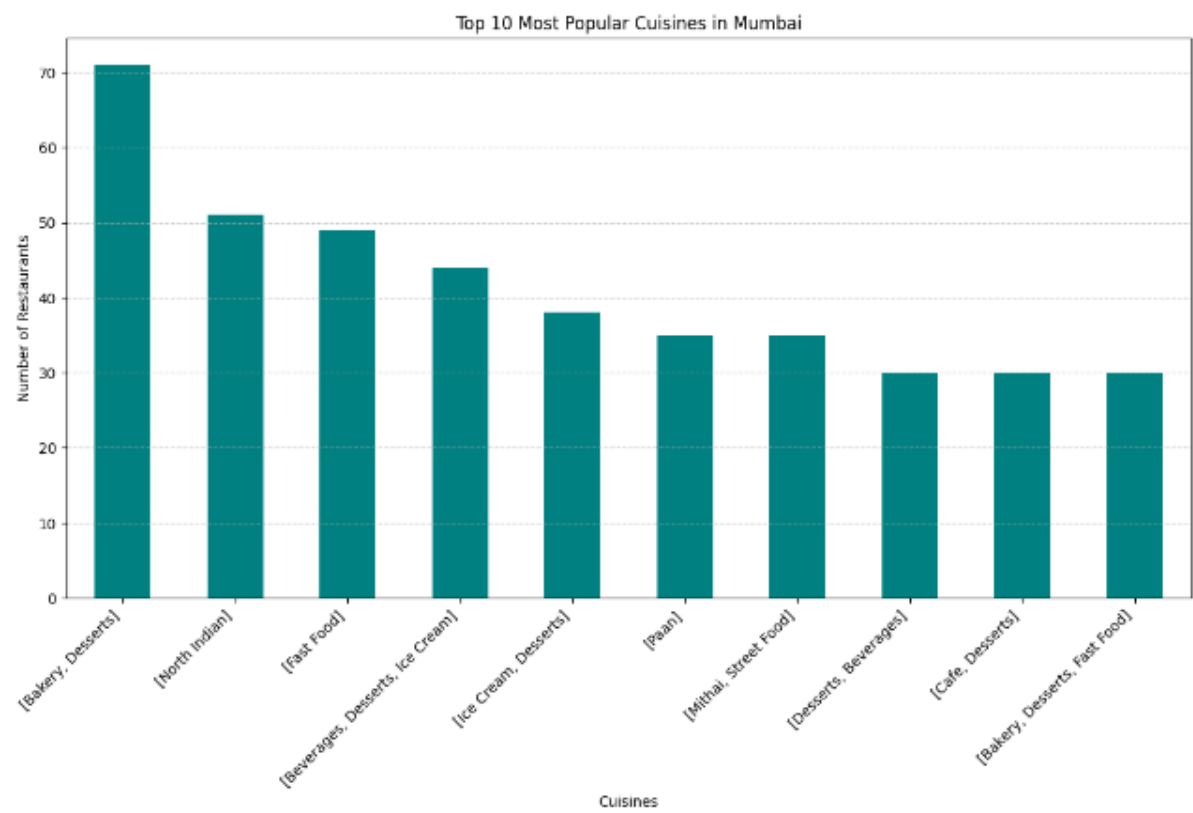- Price Range 4 (most expensive): 187 restaurants

Top Rated Restaurants:

- High Ratings: Restaurants like Biergarten, Communiti, and Asia Kitchen by Mainland China are among the highest-rated, all scoring 4.9.

Cuisine and Cost:

- Biergarten offers a mix of Continental, North Indian, and European dishes with an average cost of ₹2100 for two, positioning it in the premium segment.
- Communiti and Asia Kitchen by Mainland China offer Continental and Asian cuisines respectively, both with a mid to high price range.

Mumbai Market Analysis

Top 10 Most Popular Cuisines in Mumbai

Price Range Distribution in Mumbai Restaurants

Top Rated Restaurants in Mumbai by Average Cost for Two

Cuisine Popularity:

- Most Popular Cuisines: The top cuisines in Mumbai include North Indian, Fast Food, and a variety of dessert-related options such as Pan, Bakery, and Ice Cream.
- Diverse Offerings: There's a strong representation of both local (North Indian) and versatile options like Chinese and Continental.

Pricing Analysis:

- Average Cost for Two: The average cost for dining out in Mumbai is approximately ₹1102.
- Price Range Distribution: Most restaurants fall into the lower to mid pricing brackets, with:
- Price Range 1 (most affordable): 815 restaurants
- Price Range 2: 660 restaurants
- Price Range 3 (more premium): 659 restaurants
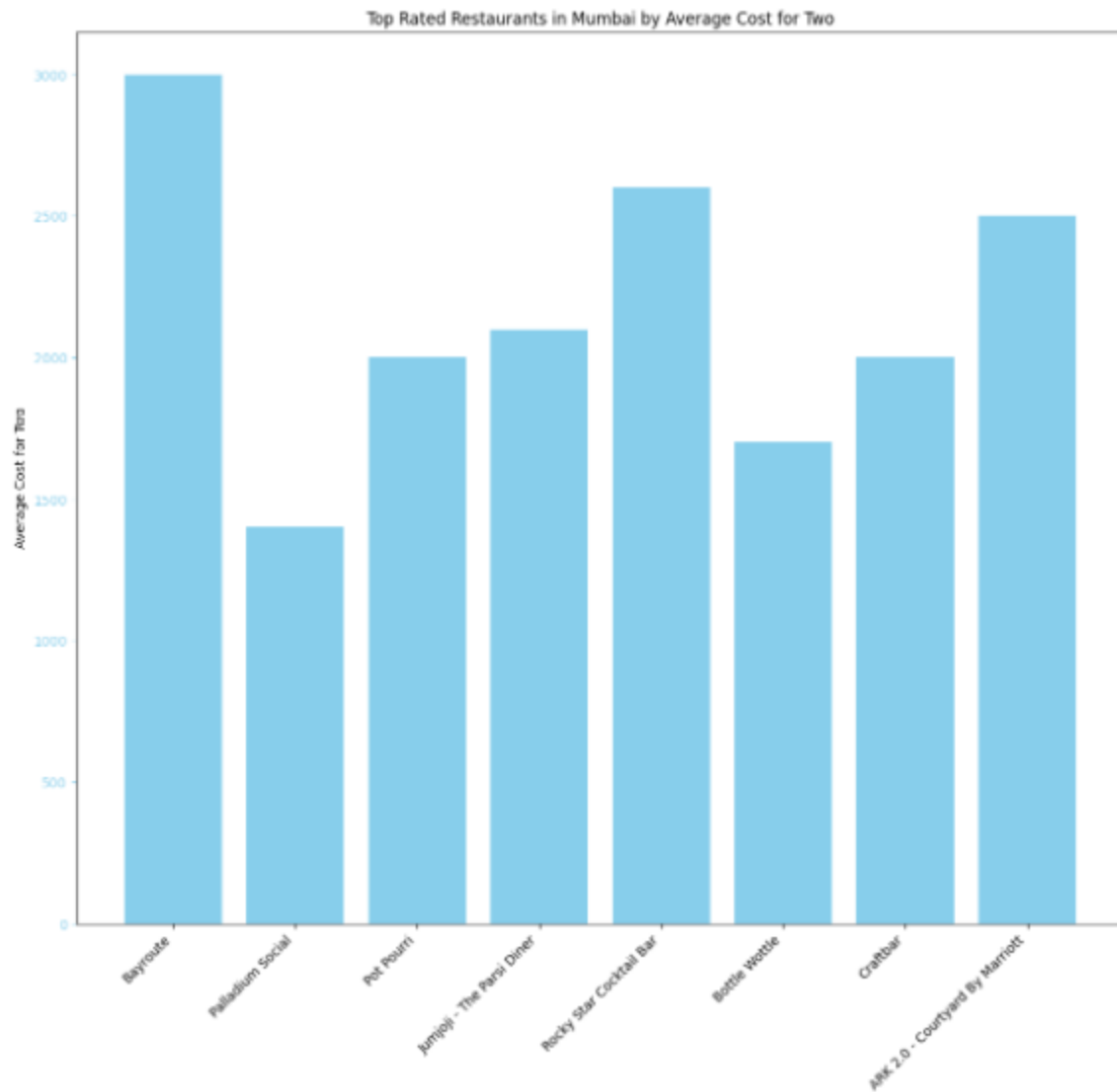- Price Range 4 (most expensive): 404 restaurants

## Top Rated Restaurants:

- High Ratings: Restaurants like Bayroute, Palladium Social, and Pot Pourri are among the highest-rated, all scoring 4.9.
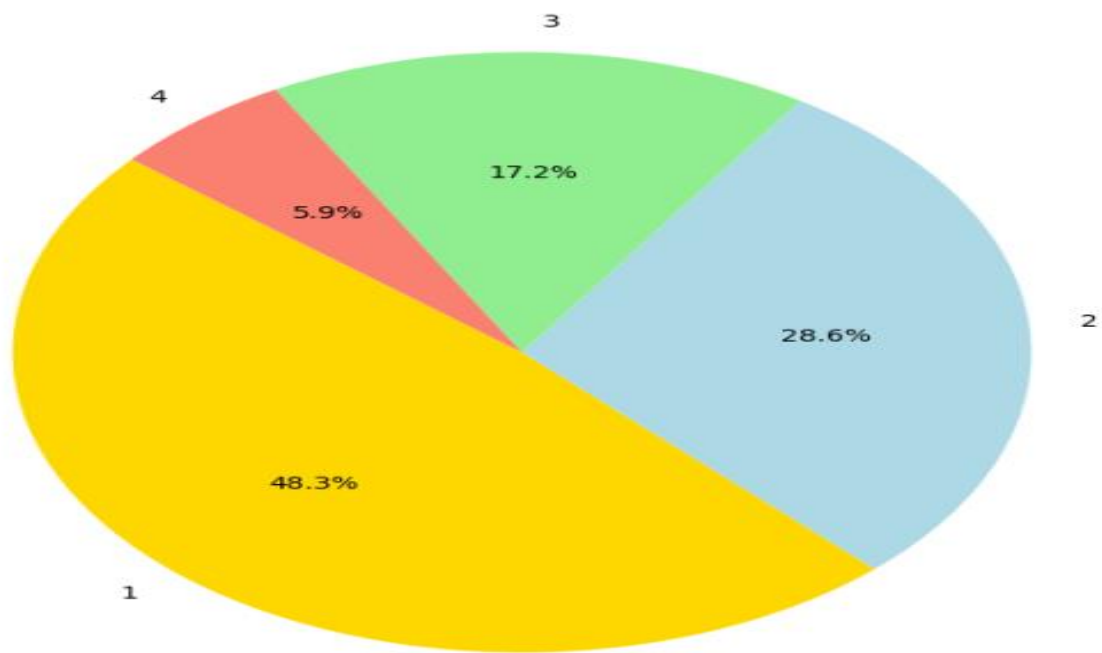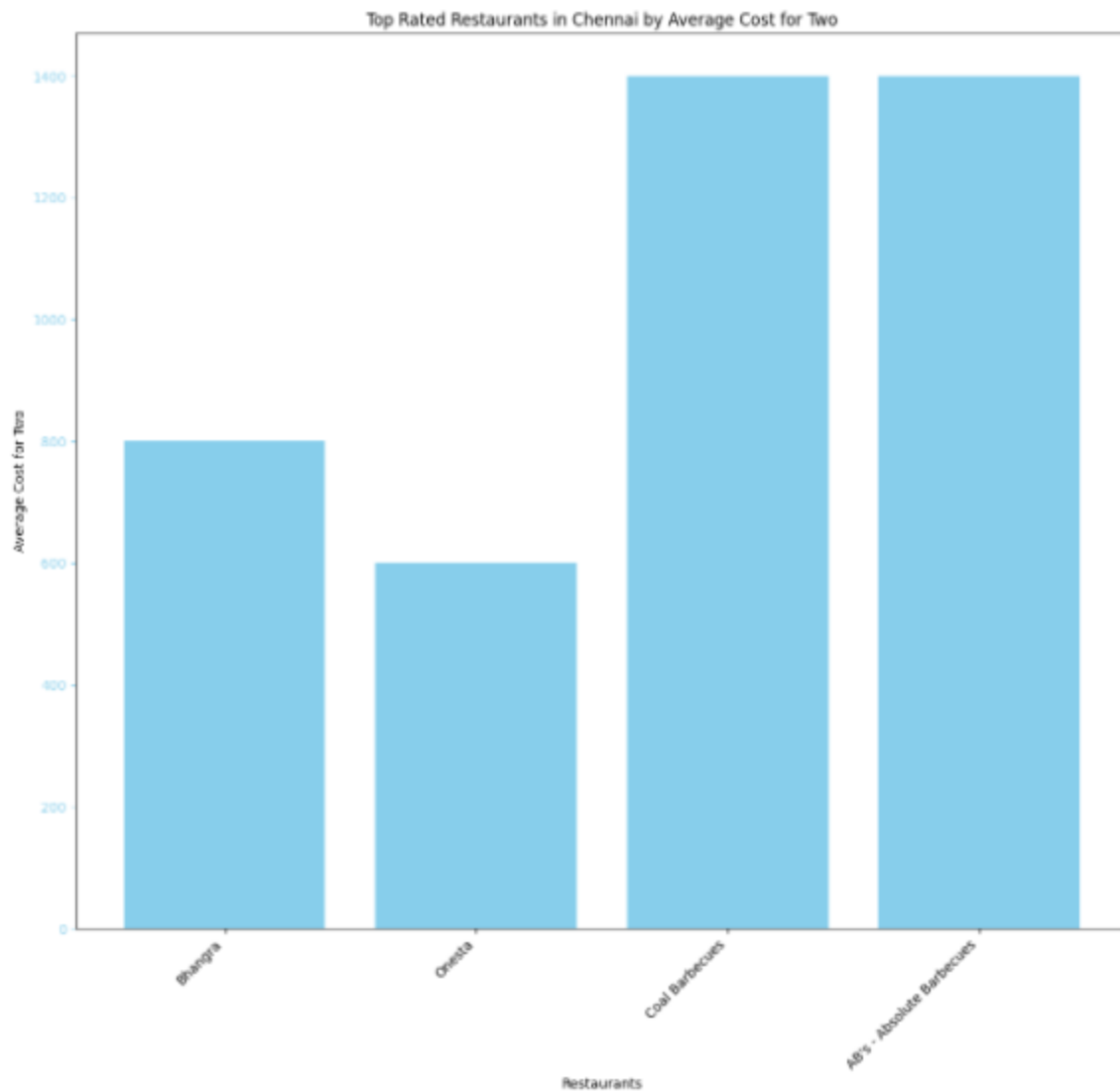
Cuisine and Cost:

- Bayroute offers a mix of Turkish and European dishes with an average cost of ₹3000 for two.
- Palladium Social and Pot Pourri offer Fast Food, Chinese and Continental respectively, both with a mid to high price range.

Chennai Market Analysis

Top 10 Most Popular Cuisines in Chennai
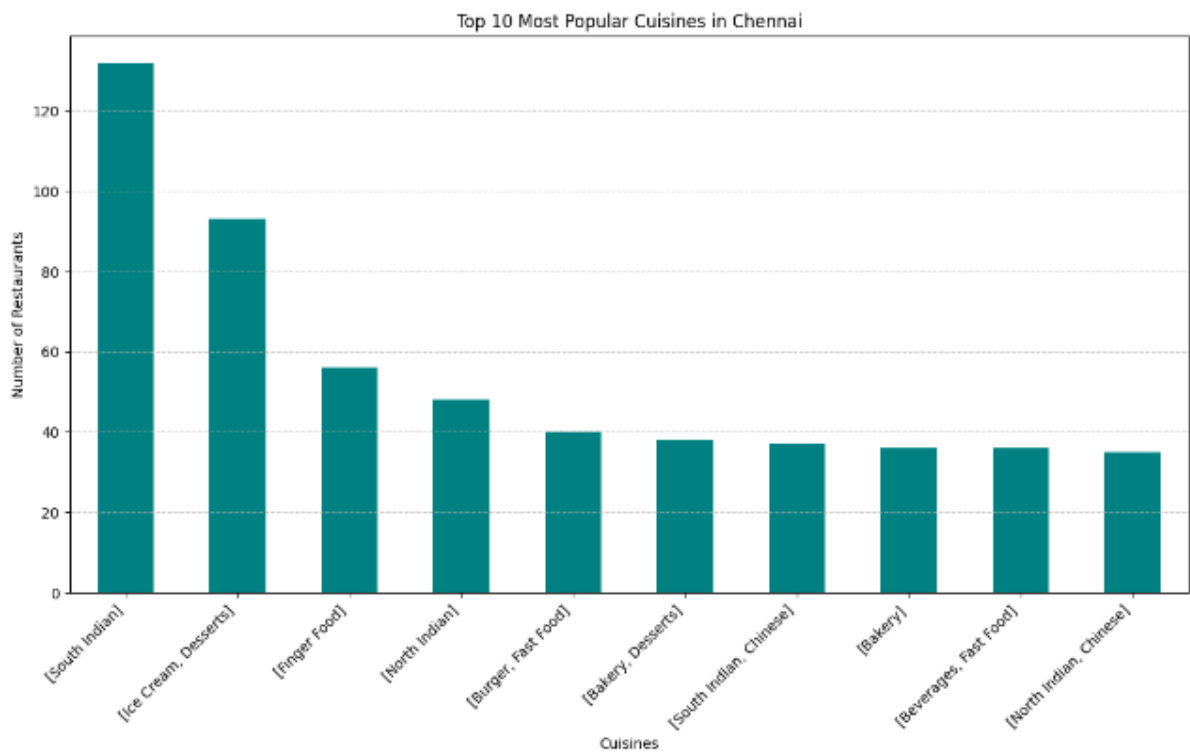
Price Range Distribution in Chennai Restaurants

Top Rated Restaurants in Chennai by Average Cost for Two

Cuisine Popularity:

- Most Popular Cuisines: The top cuisines in Bangalore include South Indian, Finger Food, North Indian, and a variety of options such as Desserts and Ice Cream.
- Diverse Offerings: There is a good mix of traditional (North Indian, South Indian) and international (Arabian, Asian) cuisines

Pricing Analysis:

- Average Cost for Two: The average cost for dining out in Bangalore is approximately ₹708.
- Price Range Distribution: Most restaurants fall into the lower to mid pricing brackets, with:
- Price Range 1 (most affordable): 1261 restaurants
- Price Range 2: 748 restaurants
- Price Range 3 (more premium): 450 restaurants
- Price Range 4 (most expensive): 153 restaurants

Top Rated Restaurants:

- High Ratings: Restaurants like Bhangra, Onesta, Coal Barbecues, are among the highest-rated, all scoring 4.9.

Cuisine and Cost:

- Bhangra offers North Indian dishes with an average cost of ₹800 for two.
- Onesta offers Fast Food and Italian with a affordable price range.
- Coal Barbecues offers Mediterranean, Asian and Arabian with mid price range.

Strengths and Weaknesses Analysis

Strengths and Weakness for Bangalore City

Strengths:

- Diverse Cuisine Offerings: Bangalore's mix of local and international cuisines caters well to its cosmopolitan population.
- High Customer Satisfaction: The presence of many high-rated restaurants indicates strong culinary quality and service.
- Affordability: The prevalence of lower price ranges shows that dining out is accessible to a wider audience.

Weaknesses:

- Competitive Market for Popular Cuisines: High competition in popular categories such as South - Indian and Street Food could challenge new entrants and existing players looking to expand.
- Limited Premium Dining Options: Fewer high-end restaurants might limit options for luxury dining experiences, possibly underserving a segment of the market.

Strengths and Weakness for Mumbai City

Strengths:

- Diverse Cuisine Offerings: Mumbai's restaurant scene is vibrant and diverse, catering to a wide range of taste preferences.
- High Customer Satisfaction: High ratings across top restaurants indicate strong culinary quality and service standards.
- Good Mix of Pricing: The presence of all price ranges from low to high allows accommodation of different economic segments.

Weaknesses:

- High Competition in Popular Categories: The dominance of certain cuisines like North Indian and Fast Food could mean higher competition, making it harder for new entrants to establish themselves.
- Cost Barrier: The higher average cost might deter budget-conscious diners, limiting the market for more affordable dining options.

Strengths and Weakness for Chennai City

Strengths:

- Diverse Cuisine Offerings: Chennai's restaurant scene is vibrant and diverse, catering to a wide range of taste preferences.
- High Customer Satisfaction: High ratings across top restaurants indicate strong culinary quality and service standards.
- Affordability: The prevalence of lower price ranges shows that dining out is accessible to a wider audience.

Weaknesses:

- Limited Cuisine Offerings: The dominance of certain cuisines like South Indian, Arabian, and Meditarian could mean higher competition, making it harder for new entrants to establish themselves.
- Limited Premium Dining Options: Lower presence in higher price ranges might indicate a lesser focus on premium dining experiences, which could be an untapped market.

Market Gap Analysis

Average Cost for Two Across Top Cities

Average Cost for Two Across Bottom Cities
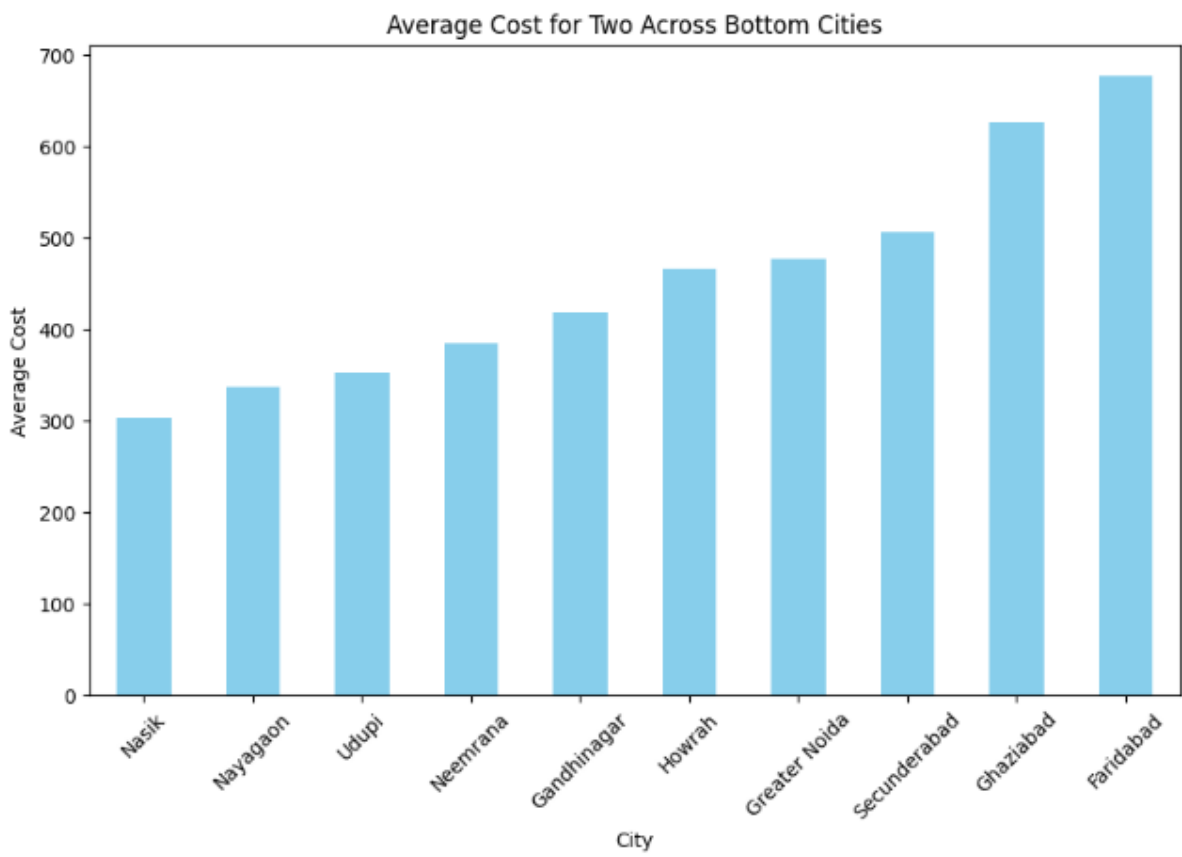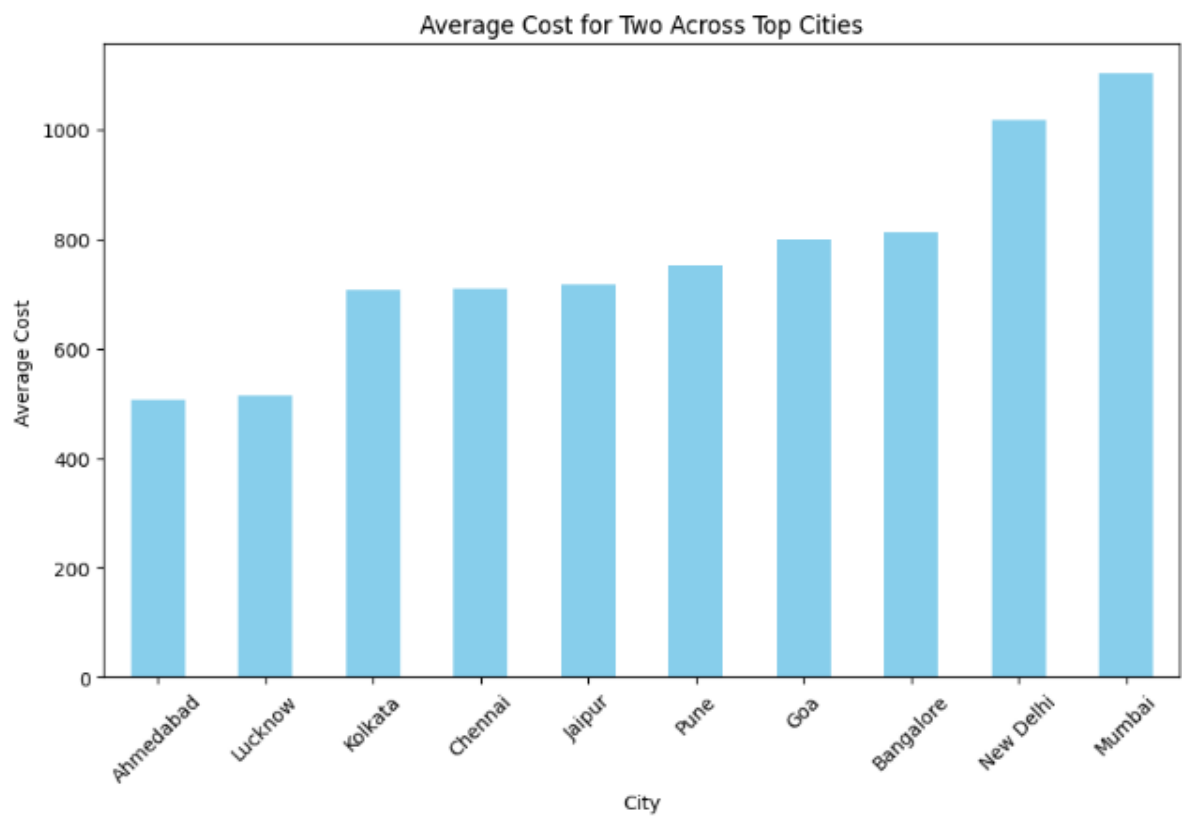
Underrepresented Cuisines with High Ratings Across All Regions

	Cuisine	Average Rating	Number of Restaurants
0	Egyptian	4.685333	6
5	Tex-Mex	4.676846	33
1	Moroccan	4.536937	12
4	Singaporean	4.481890	16
6	Greek	4.448529	22
2	Poké	4.402703	12
3	Fusion	4.370000	2
8	South American	4.282353	7
7	Mongolian	4.142466	9

Here are some cuisines that are underrepresented across different regions in our dataset, yet have received high average ratings, indicating a potential market opportunity:

- Egyptian: Very high average rating of 4.68, but only 6 restaurants offer this cuisine.
- Moroccan: Average rating of 4.53 with only 12 establishments.
- Poké: An emerging trend with an average rating of 4.4 and only 12 outlets.
- Fusion: Though only 2 outlets offer it, it has a rating of 4.37.
- Singaporean: Solid average rating of 4.48 with 16 locations.
- Tex-Mex: More commonly available with 33 locations but still relatively few given its average rating of 4.67.
- Greek: Average rating of 4.44 across 22 locations.
- Mongolian and South American: Both cuisines are scarcely represented, with 9 and 7 locations respectively, but hold respectable ratings.

Explore Niche Cuisines: Invest in cuisines such as Egyptian, Moroccan, and other culturally unique offerings, which are currently underrepresented and could differentiate a new restaurant in a competitive market.

Overall Price Range Distribution Across All Regions

Overall Price Range Distribution Across All Regions
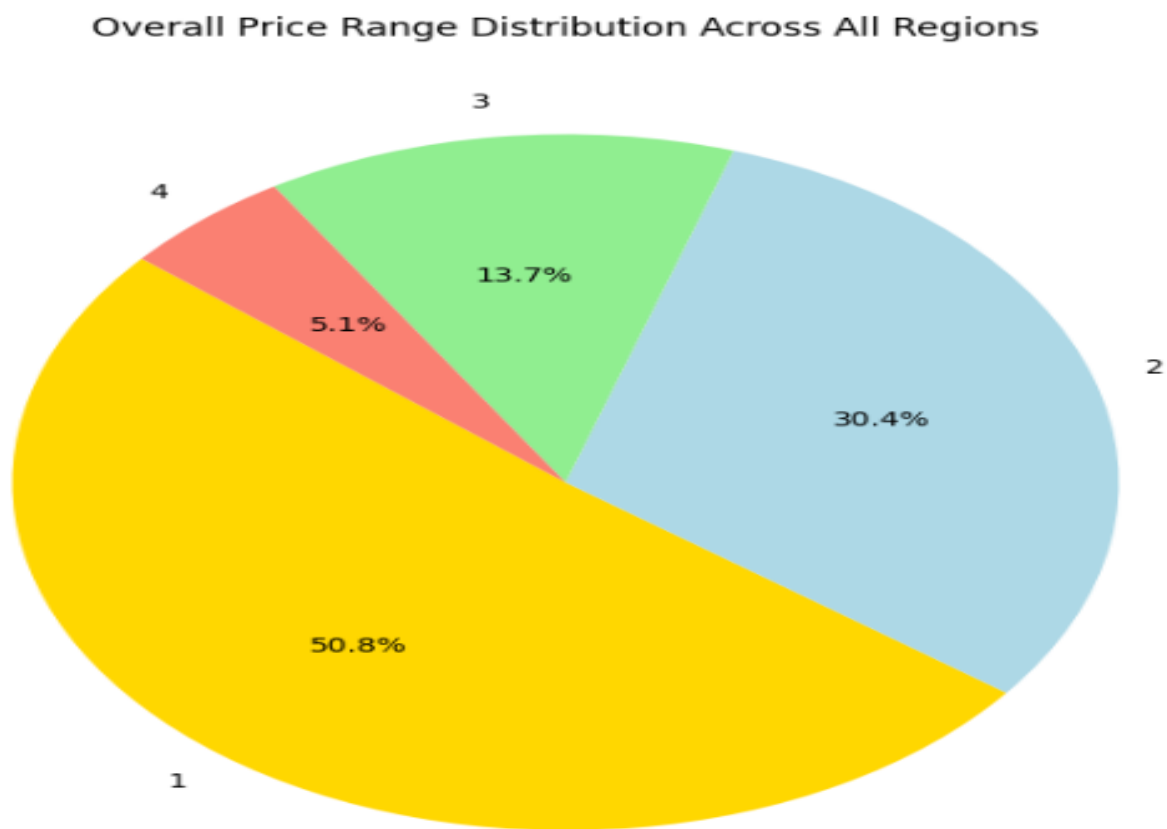
The pie chart and the data show the distribution of price ranges across all restaurants in the dataset:

- Price Range 1 (Cheapest): 30,680 restaurants (50.8%)
- Price Range 2: 18,378 restaurants (30.4%)
- Price Range 3: 8,301 restaurants (13.7%)
- Price Range 4 (Most Expensive): 3,058 restaurants (5.1%)

Insights on Market Gap Opportunities:
- The distribution reveals a significant concentration of restaurants in the lower price ranges (1 and 2), while the higher price ranges (3 and 4) are less populated.
- **High-End Market:** There appears to be room for growth in the premium dining segment (Price Range 4). Restaurants offering underrepresented yet highly-rated cuisines like Egyptian, Moroccan, or Fusion could potentially thrive in this segment.
- **Mid-Range Dining:** Price Range 3 is also less saturated compared to the lowest price ranges. Introducing unique cuisines in this price bracket might attract diners looking for a balance between quality and affordability.
- **Target Premium and Mid-Range Markets:** Consider launching restaurants in Price Ranges 3 and 4 that offer underrepresented cuisines with proven appeal based on high ratings.

Designing a Marketing Campaign

Campaign Design Based on Region Specific Insights:

Mumbai and Bangalore

Higher-end Market Opportunities:

- **Target Audience:** Young professionals and affluent families.
- **Campaign Focus:** Emphasize unique dining experiences and high-quality service. Highlight exclusive cuisines like Vietnamese or Turkish if introduced, catering to the cosmopolitan tastes of these cities.

Promotional Tactics:

- **Loyalty Programs:** Introduce a tiered loyalty program offering progressively better rewards like discounts, freebies, or exclusive event invitations.
- **Special Events:** Host themed dining nights or chef specials that showcase new or unique cuisines, enhancing the novelty appeal.

Hyderabad and Kolkata

Mid-range Market Opportunities:

- **Target Audience**: Middle-class families and corporate groups.
- **Campaign Focus**: Promote good value-for-money and family-friendly dining without compromising on the quality.

Promotional Tactics:

- **Discounts:** Offer introductory discounts or value meals during certain hours to attract families and large groups.
- **Seasonal Promotions:** Capitalize on local festivals and holidays to offer themed menus and special pricing.

General Strategies

General Strategies for All Regions

Customer Segmentation:

- **Cuisine Lovers**: Target customers seeking specific cuisines or dietary preferences with targeted ads and menu highlights.
- **Budget-Conscious Diners**: Offer combo meals and happy hour specials to attract this segment.

Promotional Tactics:

- **Social Media Campaigns:** Use platforms like Instagram and Facebook to run contests where participants can win dining vouchers by posting pictures at the restaurant or tagging friends.
- **Influencer Partnerships:** Collaborate with local food bloggers and influencers to reach broader audiences. This can include exclusive tastings or featured posts.
- **Email Marketing:** Send out monthly newsletters featuring menu updates, special events, and exclusive offers to keep the brand top-of-mind for customers.

Launch Strategy:

- **Soft Launch**: Begin with a soft launch inviting local influencers and critics to create buzz and gather initial feedback for refinement.
- **Grand Opening**: Organize a grand opening event with live music or entertainment to attract high footfall. Offer free samples or limited-time offers during the first week to draw in initial customers. By tailoring the marketing strategies to the specific characteristics and needs of each region, focus on unique selling propositions and customer engagement. The restaurant chain can effectively position itself in competitive markets and attract diverse customer segments.

Recommendations Based on Insights

Given our findings from the competitive and market gap analysis, we can develop a comprehensive marketing campaign aimed at introducing new restaurant concepts in underrepresented cuisines and price ranges.

- Introduce Unique Dining Experiences: Target underrepresented cuisines such as Egyptian, Moroccan, and Fusion in mid-range and premium segments.
- Expand in Less Saturated Markets: Focus on premium price ranges, offering a higher quality dining experience at a competitive price.
- Build Brand Awareness and Loyalty: Leverage promotional tactics to attract and retain customers.
- Food Enthusiasts and Adventurers: Target customers who are known for trying new cuisines and appreciate fine dining. Use food blogs, social media influencers, and gourmet food events to reach this segment.
- Affluent Diners: Focus on higher-income individuals for premium dining experiences. Offer exclusive memberships or VIP services that provide added value, such as priority reservations or special tasting menus.
- Discounts and Special Offers: Provide opening week discounts, off-peak hour discounts, or bundled meal deals to attract initial customers.
- Loyalty Programs: Implement a rewards program where frequent diners can earn points redeemable for discounts, free items, or exclusive experiences.
- Partnerships and Collaborations: Collaborate with local celebrities, chefs, or influencers to endorse the restaurant, creating credibility and interest.
- Pre-Launch: Build anticipation with a teaser campaign. Use social media and local influencers to generate interest and excitement.
- Launch Phase: Execute the grand opening events and begin the initial promotional offers. Monitor customer feedback closely to make quick adjustments as needed.
- Post-Launch: Evaluate the effectiveness of different tactics and refine the approach. Focus on building a community around the restaurant through events, collaborations, and consistent engagement on social media.